Discrete Applied Mathematics 157 (2009) 3552–3557

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Finding induced trees

N. Derhy, C. Picouleau *

CEDRIC, CNAM, Paris, France

ARTICLE INFO

Article history:
Received 16 November 2007
Received in revised form 26 January 2009
Accepted 17 February 2009
Available online 19 March 2009

Keywords:
Induced subgraph
Induced tree
$\mathcal{NP}$-completeness
Polynomial time algorithm

ABSTRACT

Let $G = (V(G), E(G))$ be a finite connected undirected graph and $W \subset V(G)$ a subset of vertices. We are searching for a subset $X \subset V(G)$ such that $W \subset X$ and the subgraph induced on X is a tree. $\mathcal{NP}$-completeness results and polynomial time algorithms are given assuming that the cardinality of W is fixed or not. Besides we give complexity results when X has to induce a path or when G is a weighted graph. We also consider problems where the cardinality of X has to be minimized.

1. Introduction

Detecting induced subgraphs plays a major role in combinatorial optimization both for its theoretical interest and for its numerous applications. Among a large variety of problems, those consisting in finding an induced subgraph with a specific structure, like a complete graph, a stable set, a bipartite graph or a cycle, are the subjects of a large literature. For some of these problems, max clique or max stable set for instance, we look for an induced subgraph of maximum size. For other problems, we have to decide if there exists a particular induced subgraph, like a hole, a pyramid (see [4]) or a prism (see [10]). Another kind of problem consists in finding an induced subgraph containing a prescribed set of vertices. In [3], Bienstock looks for an induced cycle passing through two prescribed vertices and in [5], Chudnovsky and Seymour study the existence of an induced tree covering three prescribed vertices. Similarly to this last problem, we focus on the problems of finding an induced tree or an induced path containing a prescribed set of vertices. We study these problems in weighted and unweighted graphs and we also consider some particular classes of graph like the bipartite graphs or the triangle-free graphs. Besides, after showing that the general case is $\mathcal{NP}$-complete, we consider the case where the number of prescribed vertices is bounded.

A variation of these problems consisting in finding a minimal tree (not necessarily induced) spanning a prescribed set of vertices is the well-known Steiner Tree Problem. We recall that this problem is $\mathcal{NP}$-hard and it is polynomial when the number of prescribed vertices is fixed.

Our paper is divided into four sections. In the first one, we give the notations and the definitions used in this paper. In the Section 2, we deal with weighted graphs. In the third one, we study the general case and show some $\mathcal{NP}$-completeness results even in particular classes of graph. In the Section 4 we focus on the case where the number of prescribed vertices is fixed. For three prescribed vertices, we prove that finding an induced path is an $\mathcal{NP}$-complete problem and give a polynomial algorithm for finding an induced tree of minimum size in a connected triangle-free graph (in this case, an induced tree always exists). When the set of prescribed vertices has a fixed cardinality k and when the shortest cycle of the graph has length at least $k+1$, we prove that there always exists an induced tree and that an induced tree of minimal size can be found in polynomial time. Finally, we quote the problems that remain unsolved in the last section.

* Corresponding address: Conservatoire National des Arts & Metie, Chaire de Recherche Operationnelle, 292 Rue Saint Martin, 75003 Paris, France.
E-mail addresses: nicolas.derhy@cnam.fr (N. Derhy), Christophe.Picouleau@cnam.fr, chp@cnam.fr (C. Picouleau).

0166-218X/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2009.02.009

2. Notation and terminology

All graphs in this paper are finite, simple and connected. If G is a graph, its vertex and edge sets are denoted $V(G)$, $E(G)$. If x is a vertex of G, its set of neighbors in G is denoted $N_G(x)$. Let $\Delta(G) = \max_{x\in V(G)}\{|N_G(x)|\}$. If $X \subset V(G)$, the subgraph with vertex set X and edge set containing all edges of G with both ends in X is called the subgraph *induced* on X. When no confusion may occur, we call X the graph induced on X. A graph G is H-free if there is no subset $X \subset V(G)$ such that the subgraph induced on X is isomorphic to H. For any $k \geq 3$, we write C_k the cycle on k vertices. Like this, a graph G is C_k-free if it does not contain C_k as an induced subgraph. In the case of a weighted graph we have both a graph G and a weight function $w : E(G) \longmapsto \mathbb{Z}$. For graph theoretical terms not defined here, the reader is referred to [2].

If $W \subset V(G)$, for any subset X such that $W \subset X \subset V(G)$, we say that the subgraph induced on X *covers* the set of *mandatory* vertices W. Alternatively we say that W is the subset of prescribed vertices.

In [3] Bienstock proves that the following problem is $\mathcal{NP}$-complete:
2-Induced-Cycle
Input: A graph G and $W = \{x, y\}$ two vertices of G.
Question: Is there a subset $X \subset V(G)$ such that $W \subset X$ and the subgraph induced on X is a cycle?
In [9] Lévêque et al. strengthen this result showing that 2-Induced-Cycle is $\mathcal{NP}$-complete even if G has maximum degree three and x, y have degree two.

In contrast in [5] Chudnovsky and Seymour prove that the following problem can be solved in polynomial time:
3-Induced-Tree
Input: A graph G and $W = \{x, y, z\}$ three vertices of G.
Question: Is there a subset $X \subset V(G)$ such that $W \subset X$ and the subgraph induced on X is a tree?

We formally define the problems we are interested in:
Induced-Tree
Input: A graph G and $W \subset V(G)$.
Question: Is there a subset $X \subset V(G)$ such that $W \subset X$ and the subgraph induced on X is a tree ?

For any $k \geq 1$, k-Induced-Tree
Input: A graph G and $W = \{x_1, \dots, x_k\}$ k vertices of G.
Question: Is there a subset $X \subset V(G)$ such that $W \subset X$ and the subgraph induced on X is a tree ?

Minimum-Induced-Tree
Input: A graph G, $W \subset V(G)$ and $B \in \mathbb{N}$.
Question: Is there a subset $X \subset V(G)$ such that $W \subset X$, the subgraph induced on X is a tree and $|X| \leq B$?

For any $k \geq 1$, Minimum-k-Induced-Tree
Input: A graph G, $W = \{x_1, \dots, x_k\}$ k vertices of G and $B \in \mathbb{N}$.
Question: Is there a subset $X \subset V(G)$ such that $W \subset X$, the subgraph induced on X is a tree and $|X| \leq B$?

Restricting the subgraph induced on X to be a path (resp. a cycle), the problems Induced-Path, k-Induced-Path, Minimum-Induced-Path, Minimum-k-Induced-Path (resp. the problems Induced-Cycle, k-Induced-Cycle, Minimum-Induced-Cycle, Minimum-k-Induced-Cycle) are defined in the same manner.

Weighted Induced-Tree
Input: A weighted graph G, $W \subset V(G)$ and $B \in \mathbb{N}$.
Question: Is there a subset $X \subset V(G)$ such that $W \subset X$, the subgraph induced on X is a tree and $\Sigma_{e\in E(X)} w(e) \leq B$?
In a similar way we define the problems Weighted k-Induced-Tree, Weighted Induced-Path, Weighted k-Induced-Path.

The problems defined above are decision problems. When no confusion occurs we call by the same name the decision problem and the problem consisting in finding the corresponding subgraph induced on X.

Recall that in a connected graph G, for any $\{x, y\} \subset V(G)$ every shortest path between x and y is chordless. It follows that 2-Induced-Path (Minimum-2-Induced-Tree and 2-Induced-Tree) can be solved in linear time using a breadth-first search (note that in contrast 2-Induced-Cycle is $\mathcal{NP}$-complete).

Finally remark that if the subgraph induced on the prescribed subset W is not acyclic, there cannot be an induced tree (or path) covering W and the problems defined above are trivial.

3. The weighted problems

Theorem 3.1. Weighted 2-Induced-Tree *and* Weighted 2-Induced-Path *are strongly $\mathcal{NP}$-complete in bipartite graphs even if the weights are* 0 *or* 1.

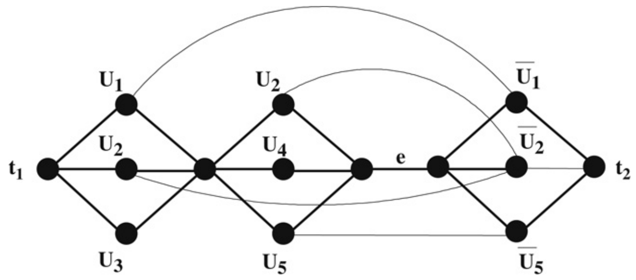

Fig. 1. The graph obtained for the instance $(u_1 \vee u_2 \vee u_3) \wedge (u_2 \vee u_4 \vee u_5) \wedge (\bar{u_1} \vee \bar{u_2} \vee \bar{u_5})$.

Proof. We use a reduction from MONOTONE 3-SAT [7]. Recall that in MONOTONE 3-SAT the three variables of a clause are either on their positive form or on their negative form. In the first case we say that the clause is positive and negative in the second case.

Let I be an instance of MONOTONE 3-SAT. We construct an instance I' of WEIGHTED 2-INDUCED-TREE by the following way (see Fig. 1). For each clause of I, we create a $K_{2,3}$ clause component. The three independent vertices forming one part of the bipartition correspond to the variables of the clause and the two other independent vertices are used to connect the clause components. We connect all the clause components corresponding to the positive clauses and do the same with the negative clause components. Finally an edge e connects the positive components to the negative components. The weight of e and of the edges of these components are equal to 0. Besides, we connect by an edge of weight 1 every pair of vertices associated with u_i and $\bar{u_i}$. Finally, the two mandatory vertices t_1 and t_2 are the two vertices not used to connect components. Note that the constructed graph is bipartite.

We claim that there is a truth assignment for I if and only if we can find a solution for I whose weight is less than or equal to 0. If there is a truth assignment for I, we construct an induced path from t_1 to t_2. In each clause component, the path uses a vertex corresponding to any true variable of this clause. Thus the weight of the path is equal to 0 and is necessarily induced because u_i and $\bar{u_i}$ cannot be true at the same time. Conversely, a solution of I' cannot use edges of weight equal to 1. Besides, it is necessarily a path because we cannot obtain an induced tree covering t_1 and t_2 which is not a path without using edges of weight 1. Since the path is induced, it cannot use a vertex corresponding to u_i and another corresponding to $\bar{u_i}$. So, we obtain a solution for I by setting to true the variables corresponding to the vertices used by the path. □

Besides, for the optimization problem, we obtain the following result:

Corollary 3.1. *Unless $\mathcal{P} = \mathcal{NP}$, there cannot exist a polynomial time approximation algorithm with a relative performance guarantee for the optimization problem associated to* WEIGHTED 2-INDUCED-TREE *even for a bipartite graph where the weights are* 0 *or* 1.

Proof. This result is a direct consequence of the reduction of Theorem 3.1 because if such an algorithm would exist, we would have a polynomial algorithm for MONOTONE 3-SAT. Indeed, we look for a solution of weight less than 0 so the approximation algorithm would find a solution of weight 0 which would give the solution of the instance of monotone 3-SAT. □

We can notice that these results can be extended to WEIGHTED K-INDUCED-TREE and WEIGHTED K-INDUCED-PATH for any fixed $k \geq 2$.

4. The unweighted problems

Since the weighted version of INDUCED-TREE and INDUCED-PATH are strongly $\mathcal{NP}$-complete even for two mandatory vertices, from now on we focus our study on the problems with unweighted graphs.

Theorem 4.1. INDUCED-TREE *and* INDUCED-PATH *are strongly $\mathcal{NP}$-complete even for planar bipartite cubic graphs.*

Proof. The reduction is from the Hamiltonian path problem which is strongly $\mathcal{NP}$-complete in planar bipartite cubic graphs (see [1]). The input of the Hamiltonian path problem consists in $H = (S, A)$ a planar bipartite cubic graph. The question is to decide if there exists an Hamiltonian path in H.

From H we build an instance of INDUCED-TREE (resp. INDUCED-PATH) as follows: each vertex $s \in S$ is replaced by the graph depicted in Fig. 2. In this graph we call x the *inner* vertex, we call u, v, w the *outer* vertices, and the three other vertices are call *internal*. If $[s, t] \in A$ is an edge of H, one outer vertex of the subgraph associated to s is connected to one outer vertex of the subgraph associated to t. The set W of mandatory vertices is the set of inner vertices. So, G is planar bipartite (cycles have an even length) and the degree of each vertex is three.

If H has a Hamiltonian path $P = (s_1, s_2, \ldots, s_n)$, then the corresponding tree (resp. path) in G (in fact a path) is obtained as shown on the right of Fig. 3. This tree (resp. path) covers W and has no chord. So, the subgraph induced by the vertices of this path is acyclic and covers W.

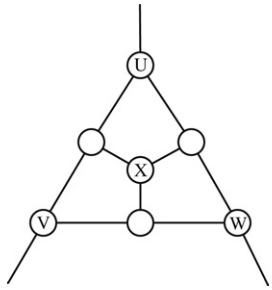

Fig. 2. The gadget replacing a vertex $s \in S$, the inner vertex x and the outer vertices u, v, w.

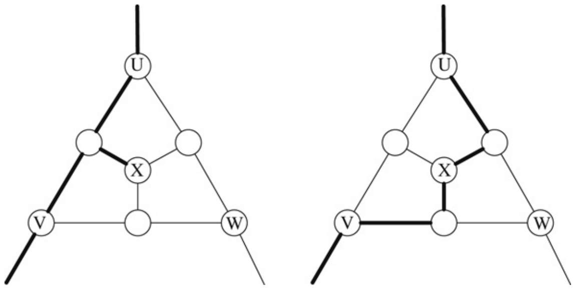

Fig. 3. The transformation of a tree into a path.

Conversely, suppose there is an induced tree covering W, say T. We give some properties of T. We can suppose that every leaf is an inner (mandatory) vertex. Since each inner vertex is a vertex of T, an outer vertex cannot be adjacent to two internal vertices (else T induces a cycle). In a gadget, the three outer vertices cannot be in T. So, if there is exactly one outer vertex, the inner vertex is a leaf of T. Otherwise, there are two outer vertices in T, say u, v. If the internal vertex adjacent to u and v is a vertex of T then, as shown on the left of Fig. 3, this internal vertex has degree three and the inner vertex is a leaf of T. Then, as shown in Fig. 3, T can be transformed into another induced tree where the degree of the inner vertex is two. Making this for all the gadgets we have an induced path P that covers W. This path corresponds to a Hamiltonian path in H. $\square$

Theorem 4.2. INDUCED-CYCLE *is strongly $\mathcal{NP}$-complete even for planar bipartite cubic graphs.*

Proof. The transformation is similar to the previous one, using a reduction from the Hamiltonian cycle problem which is $\mathcal{NP}$-complete in planar bipartite cubic graphs (see [1]). $\square$

5. Fixed number of mandatory vertices

In Section 3, we have shown that WEIGHTED INDUCED-TREE is strongly $\mathcal{NP}$-complete even for $|W| = 2$. One can remark that for both 2-INDUCED-TREE and 2-INDUCED-PATH, a solution is given by a shortest path between the two mandatory vertices since a shortest path has no chord. In this case, a solution of 2-INDUCED-TREE is also an optimal solution for the STEINER-TREE problem. Besides, in [3], Bienstock shows that the problem becomes $\mathcal{NP}$-complete if the length of the path has to be odd (the same holds if the length of the path has to be even). From now on, we focus our study on the case $|W| > 2$.

As noted in Section 2, if the mandatory subset W induces a cycle, INDUCED-TREE and INDUCED-PATH have no solution. Otherwise, in the case where G contains an edge $[u, v]$ such that $u, v \in W$ are both mandatory vertices, in order to reduce the size of the problem, we consider the graph $G' = (X', E')$ obtained from G as follows: we delete all the vertices $x \in N_G(u) \cap N_G(v)$, we contract u and v into a new vertex t and we add an edge $[t, y] \in E'$ if the vertex y is either a neighbor of u or a neighbor of v in G. Finally, we let $W' = W \cup \{t\} - \{u, v\}$ and we have:

Lemma 5.1. INDUCED-TREE *in G has a solution if and only if there is a solution for* INDUCED-TREE *in G'.*

Proof. If T is an induced tree covering W in G then $[u, v]$ is necessarily an edge of T and T cannot cover a vertex $x \in N_G(u) \cap N_G(v)$. Replacing $[u, v]$ by the vertex t, we have an induced tree T' which covers W'. Conversely, if T' is an induced tree covering W' in G', the induced graph T, obtained by replacing t by $[u, v]$ is a tree covering W. $\square$

With a similar proof we have:

Lemma 5.2. INDUCED-PATH *in G has a solution if and only if there is a solution for* INDUCED-PATH *in G'.*

5.1. Three mandatory vertices

Theorem 5.1. 3-Induced-Path *is* $\mathcal{NP}$*-complete even if* $\Delta(G) \leq 3$.

Proof. We use a reduction from 2-Induced-Cycle which is $\mathcal{NP}$-complete even if the degree of the two mandatory vertices is equal to 2 and the maximum degree is three [9]. In 2-Induced-Cycle, we are given a graph G and two mandatory vertices u, v whose degree is equal to 2. The problem is to decide whether there exists an induced cycle passing through u and v.

Let u_1 and u_2 be the two neighbors of u and consider the graph G' obtained by deleting u, by adding two new vertices u'_1 and u'_2 connected respectively to u_1 and u_2, and by setting $W = \{u'_1, u'_2, v\}$. Clearly, we have a solution of 2-Induced-Cycle in G if and only if we have a solution of 3-Induced-Path in G'. □

Theorem 5.2. *If* G *is* C_3*-free,* 3-Induced-Tree *always has a solution.*

Proof. The proof is by induction on n, the number of vertices of G. Let us denote by u, v, w the three mandatory vertices. For $n = 3$, since G is C_3-free and connected, G is a path with three vertices. So, G covers u, v and w.

Let us suppose that the theorem holds until $n - 1 \geq 3$. Let $P = \{u = p_0, \ldots, p_i = v\}$, $Q = \{u = q_0, \ldots, q_j = w\}$, $R = \{v = r_0, \ldots, r_k = w\}$ be three shortest paths respectively from u to v, from u to w and from v to w. If $|P \cup Q \cup R| < n$ then the theorem holds thanks to the induction hypothesis. Now, we assume that $|P \cup Q \cup R| = n$.

If two paths, say P and Q, are such that $P \cup Q$ is an induced tree then $P \cup Q$ is the desired subgraph covering u, v and w. Now we study the case where there is a chord in every induced subgraph $P \cup Q$, $P \cup R$ and $Q \cup R$. Let p_s be the vertex of P such that there is a chord $[p_s, q_t]$ between P and Q or a chord $[p_s, r_z]$ between P and R (by symmetry, we will only consider the first case) and such that s is minimum. Since G is C_3-free, we cannot have $s = t = 1$. So by symmetry, we assume that $t > 1$. Let $P'' = \{u = p_0, \ldots, p_s\}$, $P' = \{p_s, \ldots, p_i = v\}$ and $Q' = \{q_t \ldots, q_j = w\}$ be the subpaths of P and Q demarcated by the chord $[p_s, q_t]$. Let G' be the graph induced by the vertices $P' \cup Q' \cup R$. G' is connected and by the induction hypothesis there exists an induced tree T' covering v, w and p_s. Now, from the minimality of s, we obtain that $T = P'' \cup T'$ is an induced tree covering u, v, w. This concludes the proof. □

Theorem 5.3. *If* G *is* C_3*-free,* Minimum-3-Induced-Tree *can be solved in linear time.*

Proof. Let us denote by $U(x)$, $V(x)$ and $W(x)$ the length of a shortest path from a vertex x to u, v and w respectively. Let s be a vertex such that $U(s) + V(s) + W(s) = \min_{x \in V(G)}\{U(x) + V(x) + W(x)\}$.

Let P_{us} (resp. P_{vs}, P_{ws}) be a shortest path from u (resp. v, w) to s. Let us consider the graph H induced by $P_{us} \cup P_{vs} \cup P_{ws}$. We show that H is a tree: since P_{us} (resp. P_{vs}, P_{ws}) is a shortest path it cannot have a chord; now, suppose that H has an edge $[a, b]$ with $a \in P_{us}$ and $b \in P_{vs}$. Since G is C_3-free, a and b cannot be adjacent to s and we can suppose that $S(a)$, the length of the path $P_{as} \subseteq P_{us}$, is at least two. Denoting by $S(b)$ the length of the path $P_{bs} \subseteq P_{vs}$, we have three paths starting from b, one to v of length $V(s) - S(b)$, one to u of length $U(s) - S(a) + 1$, one to w of length $W(s) + S(b)$, and we obtain $V(s) - S(b) + U(s) - S(a) + 1 + S(b) + W(s) < U(s) + V(s) + W(s)$, a contradiction.

Three breadth-first searches from u, v, w give respectively $U(x)$, $V(x)$, $W(x)$ (and their corresponding paths). Thus a minimum induced tree covering u, v, w can be computed in $O(m)$. □

5.2. k mandatory vertices

Lemma 5.3. *If* G *is* $C_{k'}$*-free for* $k' \in \{3, \ldots, k\}$*, a tree of minimal size covering* k *mandatory vertices (i.e. a minimum Steiner tree) is an induced tree.*

Proof. Let T be a tree of minimal size covering W, a set of k mandatory vertices. T being minimal, each leaf of T is a vertex of W.

Suppose that T induces a chord e. Thus the graph $T' = (V(T), E(T) \cup \{e\})$ has exactly one cycle C whose length is necessarily greater than k.

Suppose that there is no vertex of $C \setminus W$ of degree two in T. Since the length of C is greater than k and since each leaf of T is a vertex of W, there would be at least $k + 1$ mandatory vertices. So, there exists a vertex $v_C \in C \setminus W$ with degree two. Let e_1 and e_2 be the two edges of C which are incident to v_C. The graph $T'' = (V(T) \setminus \{v_C\}, E(T) \setminus \{e_1, e_2\} \cup \{e\})$ is a tree covering W and its size is less than the size of T which is minimal, a contradiction. □

Theorem 5.4. *If* G *is* $C_{k'}$*-free for* $k' \in \{3, \ldots, k\}$*,* k-Induced-Tree *always has a solution and* Minimum-k-Induced-Tree *can be solved in polynomial time.*

Proof. It is a direct consequence of Lemma 5.3: G being connected, a Steiner Tree always exists and a minimum Steiner Tree can be computed in polynomial time when $|W| = k$ is fixed (see [6]). A polynomial algorithm with time-complexity $O(3^k n + 2^k m \cdot log(n))$ is given in [8]. □

Note that for weighted graphs, Lemma 5.3 cannot be applied and the problem becomes $\mathcal{NP}$-complete (see Theorem 3.1). Besides, for $k = 3$, the time-complexity of the current best known algorithm for the Steiner Tree problem (when $|W|$ is fixed) is worse than the time-complexity of the algorithm designed in Theorem 5.3.

6. Conclusion

We studied problems consisting in searching an induced tree covering a prescribed set of vertices. We proved the $\mathcal{NP}$-completeness in the general case for both weighted and unweighted graphs. In the case where the set to be covered has a fixed cardinality k and the shortest cycle of the graph has length at least $k+1$, we proved that there is always a solution and one of minimal size can be found in polynomial time. For $k = 3$ and when the graph has no triangle, a linear time algorithm that computes an induced tree of minimum size is furnished.

Some questions are left open: in our opinion, the most challenging is the complexity status of the problem consisting in finding an induced tree covering a prescribed set of four vertices. We also mention the problem consisting in finding an induced tree of minimum size containing three prescribed vertices.

Acknowledgments

The authors are grateful to Nicolas Trotignon for some helpful discussions.

We are also grateful to the two anonymous referees for their helpful comments which helped in improving the quality of this article.

References

[1] T. Akiyama, T. Nishizeki, N. Saito, NP-completeness of the Hamilton cycle problem in bipartite graphs, J. Inf. Process. 5 (1980) 73–76.
[2] C. Berge, Graphes, Gauthier-Villars, Paris, 1983.
[3] D. Bienstock, On the complexity of testing for even holes and induced odd paths, Discrete. Math. 90 (1991) 85–92. Corrigendum in Discrete Math. 102 (1992), 109.
[4] M. Chudnovsky, G. Conuéjols, X. Liu, P. Seymour, K. Vušković, Recognizing Berge Graphs, Combinatorica 25 (2005) 143–186.
[5] M. Chudnovsky, P. Seymour, The three-in-a-tree problem, 2006 (submitted for publication).
[6] S.E. Dreyfus, R.A. Wagner, The Steiner problem in graphs, Networks 1 (1972) 195–207.
[7] M.R. Garey, D.S. Johnson, Computers and Intractability, a Guide to the Theory of $\mathcal{NP}$-Completeness, Freeman, New York, 1979.
[8] B. Kimelfeld, Y. Sagiv, New algorithms for computing Steiner trees for a fixed number of terminals, unpublished manuscript.
[9] B. Lévêque, D. Lin, F. Maffray, N. Trotignon, Detecting induced subgraphs, Discrete Appl. Math. 157 (2009) 3540–3551 (special issue GO VI).
[10] F. Maffray, N. Trotignon, Algorithms for perfectly contractile graphs, SIAM J. Discrete Math. 19 (2005) 553–574.